

## Supplementary Materials

### Section S1. Selected features after Spearman correlation filter and Wilcoxon rank-sum test steps of feature selection.

<i>Feature Type</i>	<i>Feature name</i>
Shape [n = 14]	shape Flatness shape Sphericity Maximum2D Diameter Row shape MeshVolume shape Elongation shape Maximum2DDiameterColumn shape VoxelVolume shape SurfaceArea shape MinorAxisLength shape Maximum3DDiameter shape MajorAxisLength shape LeastAxisLength shape SurfaceVolumeRatio shape Maximum2DDiameterSlice
First Order [n = 18]	Skewness Maximum MeanAbsoluteDeviation Kurtosis Range Energy InterquartileRange Variance Mean 10Percentile 90Percentile Uniformity Median RobustMeanAbsoluteDeviation RootMeanSquared Minimum TotalEnergy Entropy
Second order (GLCM) [n = 24]	glcm SumSquares glcm Contrast glcm InverseVariance glcm ClusterTendency glcm DifferenceAverage glcm Idm glcm ClusterShade glcm Imc2 glcm Idn glcm ClusterProminence glcm Id glcm Imc1 glcm Autocorrelation glcm SumEntropy glcm MaximumProbability glcm MCC glcm Correlation glcm JointEntropy

	glcm Idmn glcm JointAverage glcm SumAverage glcm DifferenceEntropy glcm DifferenceVariance glcm JointEnergy
<b>Second Order (GLRLM)</b> [n = 16]	glrlm ShortRunHighGrayLevelEmphasis glrlm RunLengthNonUniformity glrlm ShortRunEmphasis glrlm LongRunLowGrayLevelEmphasis glrlm RunPercentage glrlm GrayLevelVariance glrlm LowGrayLevelRunEmphasis glrlm GrayLevelNonUniformity glrlm RunVariance glrlm LongRunHighGrayLevelEmphasis glrlm ShortRunLowGrayLevelEmphasis glrlm RunLengthNonUniformityNormalized glrlm LongRunEmphasis glrlm GrayLevelNonUniformityNormalized glrlm HighGrayLevelRunEmphasis glrlm RunEntropy
<b>Second Order (GLSZM)</b> [n = 16]	glszm SizeZoneNonUniformity glszm LargeAreaEmphasis glszm SmallAreaEmphasis glszm ZonePercentage glszm GrayLevelVariance glszm LargeAreaLowGrayLevelEmphasis glszm GrayLevelNonUniformity glszm SmallAreaHighGrayLevelEmphasis glszm ZoneVariance glszm LargeAreaHighGrayLevelEmphasis glszm HighGrayLevelZoneEmphasis glszm SizeZoneNonUniformityNormalized glszm ZoneEntropy glszm SmallAreaLowGrayLevelEmphasis glszm GrayLevelNonUniformityNormalized glszm LowGrayLevelZoneEmphasis
<b>Second Order (NGTDM)</b> [n = 5]	ngtdm Coarseness ngtdm Strength ngtdm Busyness ngtdm Complexity ngtdm Contrast
<b>Second Order (GLDM)</b> [n = 14]	gldm LargeDependenceLowGrayLevelEmphasis gldm DependenceNonUniformity gldm SmallDependenceLowGrayLevelEmphasis gldm LowGrayLevelEmphasis gldm LargeDependenceHighGrayLevelEmphasis gldm DependenceEntropy gldm DependenceVariance gldm GrayLevelVariance gldm GrayLevelNonUniformity gldm SmallDependenceEmphasis gldm LargeDependenceEmphasis

	gldm SmallDependenceHighGrayLevelEmphasis gldm HighGrayLevelEmphasis gldm DependenceNonUniformityNormalized
--	--

**Table S1.** Extracted radiomic features. First (n=18) and second order features (n = 75) were extracted from T2 and arterial, portal and tardive post-contrast phases of DCE-MRI images, for a total of 93 features for each of the four MRI images. For each VOI, a total of 386 radiomic features were extracted.

## Section S2. PyRadiomics parameter file

#Brancato et al. - Parameter file for feature extraction with PyRadiomics - August 2021

# MRI-based radiomic features help identify lesions and predict histopathological grade of hepatocellular carcinoma

```
imageType:
  Original: {}
```

```
featureClass:
  shape:
  firstorder:
  glcm:
  glrlm:
  glszm:
  gldm:
  ngtdm:
```

```
setting:
  # Normalization:
  normalize: true
  normalizeScale: 100 # This allows you to use more or less the same bin width.

  # Image discretization:
  # The ideal number of bins is somewhere in the order of 16-128 bins. Binwidth
  was defined by extracting firstorder:Range from the dataset to analyze, and
  choosing a binwidth so, that range/binwidth remains approximately
  # in this range of bins (applied on T2, ART, PORT, TARD images).
  binWidth = 6

  # first order specific settings:
  # When normalizing, gray values below the mean will be negative. Shifting by
  300 (3 StdDevs * 100) ensures that the
  # majority of voxels is positive (only outliers >3 SD lower than the mean will
  be negative).
  voxelArrayShift: 300

  # Misc:
  # label value
  label: 1
```

### Section S3. Selected features after Spearman correlation filter and Wilcoxon rank-sum test steps of feature selection.

SHAPE shape Sphericity
SHAPE shape Elongation
ART glcm Cluster Shade
ART glrlm Long Run High Gray Level Emphasis
ART gldm Large Dependence Low Gray Level Emphasis
ART glszm Small Area Low Gray Level Emphasis
ART firstorder Minimum
ART firstorder Kurtosis
ART firstorder Skewness
PORT gldm Large Dependence Low Gray Level Emphasis
PORT glszm SmallAreaEmphasis
PORT glszm SmallAreaLowGrayLevelEmphasis
PORT firstorder Kurtosis
TARD glrlm RunVariance
TARD glszm SmallAreaEmphasis
TARD firstorder 10Percentile
TARD firstorder Kurtosis
T2 glcm Idn
T2 glcm MCC
T2 glcm Imc1
T2 glrlm ShortRunEmphasis
T2 glrlm LongRunHighGrayLevelEmphasis
T2 gldm DependenceNonUniformityNormalized
T2 glszm SizeZoneNonUniformityNormalized
T2 glszm SmallAreaLowGrayLevelEmphasis
T2 firstorder Minimum
ART ngtdm Strength
TARD ngtdm Strength
T2 ngtdm Strength

**Table S2.** 29 selected features after Wilcoxon rank-sum test for HCC/HT classification task.

SHAPE shape Sphericity
SHAPE shape Elongation
ART glcm ClusterShade
ART glrlm LongRunHighGrayLevelEmphasis
ART gldm LargeDependenceLowGrayLevelEmphasis
ART glszm SizeZoneNonUniformityNormalized
ART glszm SmallAreaLowGrayLevelEmphasis
ART firstorder Minimum
ART firstorder Kurtosis
ART firstorder Skewness
PORT gldm LargeDependenceLowGrayLevelEmphasis
PORT glszm SizeZoneNonUniformityNormalized
PORT glszm SmallAreaEmphasis
PORT glszm SmallAreaLowGrayLevelEmphasis

PORT firstorder Kurtosis
TARD glrlm RunVariance
TARD glszm SmallAreaEmphasis
TARD firstorder 10Percentile
TARD firstorder Kurtosis
T2 glcm Idn
T2 glcm MCC
T2 glcm Imc1
T2 glrlm ShortRunEmphasis
T2 glrlm LongRunHighGrayLevelEmphasis
T2 gldm LargeDependenceLowGrayLevelEmphasis
T2 gldm DependenceNonUniformityNormalized
T2 glszm SizeZoneNonUniformityNormalized
T2 glszm SmallAreaEmphasis
T2 glszm SmallAreaLowGrayLevelEmphasis
T2 firstorder 10Percentile
T2 firstorder Minimum
ART ngtdm Strength
TARD ngtdm Strength
T2 ngtdm Strength

**Table S3.** 34 selected features after paired Wilcoxon signed-rank test for HCC/HT classification task.

ART glszm SizeZoneNonUniformityNormalized
PORT glcm MaximumProbability
PORT gldm LargeDependenceLowGrayLevelEmphasis
PORT glszm SmallAreaLowGrayLevelEmphasis
TARD glszm SizeZoneNonUniformityNormalized
T2 glszm LowGrayLevelZoneEmphasis

**Table S4.** 6 selected features after Wilcoxon rank-sum test for G1+G2/G3 classification task.

ART firstorder 10Percentile
ART firstorder Skewness
PORT firstorder Skewness
TARD firstorder Maximum
T2 gldm LowGrayLevelEmphasis
PORT ngtdm Strength

**Table S5.** 6 selected features after Wilcoxon rank-sum test for G1/G2 classification task.

SHAPE shape SurfaceVolumeRatio
ART glcm ClusterShade
ART firstorder Skewness
PORT glcm MaximumProbability
TARD gldm LargeDependenceLowGrayLevelEmphasis
T2 gldm LargeDependenceHighGrayLevelEmphasis

**Table S6.** 6 selected features after Wilcoxon rank-sum test for G1/G3 classification task.

PORT glcm MaximumProbability
PORT gldm LargeDependenceLowGrayLevelEmphasis
PORT glszm ZonePercentage
PORT glszm LargeAreaLowGrayLevelEmphasis

TARD glrlm LongRunLowGrayLevelEmphasis
PORT ngtdm Complexity

**Table S7.** 6 selected features after Wilcoxon rank-sum test for G2/G3 classification task.

## Section S4. Prediction performances of multivariable logistic regression models.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	T2 gldm DNUN	0.943 $\pm$ 0.002	0.895 $\pm$ 0.004	0.9 $\pm$ 0.003	0.897 $\pm$ 0.002	0.695 $\pm$ 0.005
2	T2 gldm DNUN T2 glrlm LRHGLE	0.958 $\pm$ 0.001	0.939 $\pm$ 0.003	0.905 $\pm$ 0.002	0.919 $\pm$ 0.002	0.768 $\pm$ 0.004
3	T2 gldm DNUN T2 glrlm LRHGLE ART firstorder Minimum	0.957 $\pm$ 0.001	0.926 $\pm$ 0.004	0.903 $\pm$ 0.002	0.913 $\pm$ 0.002	0.731 $\pm$ 0.005
4	T2 gldm DNUN T2 glrlm LRHGLE ART firstorder Minimum ART gldm LDLGLE	0.953 $\pm$ 0.002	0.907 $\pm$ 0.005	0.917 $\pm$ 0.003	0.914 $\pm$ 0.002	0.733 $\pm$ 0.005
5	T2 gldm DNUN T2 glrlm LRHGLE ART firstorder Minimum ART gldm LDLGLE T2 glszm SALGLE	0.949 $\pm$ 0.002	0.899 $\pm$ 0.005	0.903 $\pm$ 0.003	0.903 $\pm$ 0.002	0.739 $\pm$ 0.006

**Table S8.** Results of multivariate analysis for HT/HCC classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; TARD = features extracted from tardive post-contrast phase of DCE-MRI; DNNU = Dependence Non Uniformity Normalized; LRHGLE = Long Run High Gray Level Emphasis; LDLGLE = Large Dependence Low Gray Level Emphasis; SALGLE = Small Area Low Gray Level Emphasis.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	PORT glcm MP	0.743 $\pm$ 0.006	0.668 $\pm$ 0.015	0.564 $\pm$ 0.013	0.605 $\pm$ 0.006	0.224 $\pm$ 0.02
2	ART glszm SZNUN T2 glszm LGLZE	0.75 $\pm$ 0.006	0.644 $\pm$ 0.011	0.647 $\pm$ 0.009	0.643 $\pm$ 0.005	0.263 $\pm$ 0.017
3	ART glszm SZNUN T2 glszm LGLZE PORT glcm MP	0.733 $\pm$ 0.006	0.638 $\pm$ 0.011	0.677 $\pm$ 0.009	0.66 $\pm$ 0.005	0.3 $\pm$ 0.016
4	PORT gldm LDLGLE PORT glcm MP T2 glszm LGLZE ART glszm SZNUN	0.743 $\pm$ 0.007	0.638 $\pm$ 0.012	0.688 $\pm$ 0.009	0.668 $\pm$ 0.005	0.278 $\pm$ 0.02
5	PORT gldm LDLGLE PORT glcm MP T2 glszm LGLZE ART glszm SZNUN PORT glszm SALGLE	0.717 $\pm$ 0.007	0.633 $\pm$ 0.012	0.69 $\pm$ 0.009	0.669 $\pm$ 0.005	0.234 $\pm$ 0.047

**Table S9.** Results of multivariate analysis for G1+G2/G3 classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all

bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; TARD = features extracted from tardive post-contrast phase of DCE-MRI; LDLGLE = Large Dependence Low Gray Level Emphasis; MP = Maximum Probability; LGLZE = Low Gray Level Zone Emphasis; SZNUN = Size Zone Non Uniformity Normalized; SALGLE = Small Area Low Gray Level Emphasis.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	PORT ngtdm Strength	0.909 $\pm$ 0.004	0.83 $\pm$ 0.006	0.665 $\pm$ 0.015	0.79 $\pm$ 0.005	0.468 $\pm$ 0.018
2	T2 gldm LGLE ART firstorder 10Percentile	0.959 $\pm$ 0.004	0.894 $\pm$ 0.007	0.736 $\pm$ 0.018	0.862 $\pm$ 0.005	0.631 $\pm$ 0.018
3	PORT ngtdm Strength T2 gldm LGLE ART firstorder 10Percentile	0.942 $\pm$ 0.006	0.916 $\pm$ 0.005	0.717 $\pm$ 0.022	0.879 $\pm$ 0.005	0.617 $\pm$ 0.023
4	PORT ngtdm Strength T2 gldm LGLE ART firstorder 10Percentile TARD firstorder Maximum	0.939 $\pm$ 0.006	0.9 $\pm$ 0.007	0.807 $\pm$ 0.018	0.884 $\pm$ 0.006	0.664 $\pm$ 0.021
5	PORT ngtdm Strength T2 gldm LGLE ART firstorder 10Percentile TARD firstorder Maximum ART firstorder Skewness	0.919 $\pm$ 0.007	0.88 $\pm$ 0.007	0.798 $\pm$ 0.018	0.865 $\pm$ 0.006	0.609 $\pm$ 0.021

**Table S10.** Results of multivariate analysis for G1/G2 classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; TARD = features extracted from tardive post-contrast phase of DCE-MRI; LGLE = Low Gray Level Emphasis.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	ART firstorder Skewness	0.858 $\pm$ 0.005	0.826 $\pm$ 0.007	0.508 $\pm$ 0.016	0.747 $\pm$ 0.005	0.264 $\pm$ 0.05
2	PORT glcm MP ART firstorder Skewness	0.885 $\pm$ 0.005	0.826 $\pm$ 0.006	0.7 $\pm$ 0.016	0.799 $\pm$ 0.004	0.489 $\pm$ 0.016
3	T2 gldm LDHGLE ART firstorder Skewness PORT glcm MP	0.886 $\pm$ 0.007	0.843 $\pm$ 0.009	0.755 $\pm$ 0.015	0.828 $\pm$ 0.006	0.583 $\pm$ 0.016
4	T2 gldm LDHGLE ART firstorder Skewness PORT glcm MP ART glcm CS	0.851 $\pm$ 0.008	0.783 $\pm$ 0.01	0.745 $\pm$ 0.016	0.782 $\pm$ 0.007	0.484 $\pm$ 0.02
5	ART firstorder Skewness PORT glcm MP T2 gldm LDHGLE ART glcm CS SHAPE shape SVR	0.812 $\pm$ 0.009	0.741 $\pm$ 0.011	0.702 $\pm$ 0.017	0.742 $\pm$ 0.008	0.354 $\pm$ 0.025

**Table S11.** Results of multivariate analysis for G1 /G3 classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-

contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; MP = Maximum Probability; SVR = Surface Volume Ratio; LDHGLE = Large Dependence High Gray Level Emphasis; CS = Cluster Shade.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	PORT ngtdm Complexity	0.707 $\pm$ 0.006	0.569 $\pm$ 0.01	0.661 $\pm$ 0.012	0.609 $\pm$ 0.005	0.165 $\pm$ 0.06
2	PORT ngtdm Complexity PORT glszm LALGLE	0.709 $\pm$ 0.006	0.609 $\pm$ 0.01	0.664 $\pm$ 0.013	0.63 $\pm$ 0.006	0.237 $\pm$ 0.026
3	PORT ngtdm Complexity PORT glszm LALGLE TARD glrlm LRLGLE	0.693 $\pm$ 0.007	0.605 $\pm$ 0.009	0.674 $\pm$ 0.012	0.638 $\pm$ 0.005	0.215 $\pm$ 0.029
4	PORT ngtdm Complexity PORT glszm LALGLE TARD glrlm LRLGLE PORT gldm LDLGLE	0.663 $\pm$ 0.006	0.613 $\pm$ 0.009	0.639 $\pm$ 0.013	0.628 $\pm$ 0.006	0.187 $\pm$ 0.03
5	PORT gldm LDLGLE PORT ngtdm Complexity PORT glszm ZP TARD glrlm LRLGLE PORT glszm LALGLE	0.641 $\pm$ 0.007	0.616 $\pm$ 0.01	0.598 $\pm$ 0.013	0.614 $\pm$ 0.006	0.103 $\pm$ 0.033

**Table S12.** Results of multivariate analysis for G2/G3 classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; LALGLE = Large Area Low Gray Level Emphasis; LRLGLE = Long Run Low Gray Level Emphasis; ZP = Zone Percentage; LDLGLE = Large Dependence Low Gray Level Emphasis.

## Section S5. Results of DeLong test with Bonferroni correction.

TASK – HCC vs HT		
Models	p	95% CI
1-2	0.303	0-0.12
1-3	0.393	0-0.114
1-4	0.282	0-0.114
1-5	0.244	0-0.032
2-3	0.34	0-0.034
2-4	0.372	0-0.017
2-5	0.346	0-0.18
3-4	0.4	0-0.058
3-5	0.271	0-0.141
4-5	0.362	0-0.033
TASK – G1+G2 vs G3		
Models	p	95% CI
1-2	0.762	0-0.16
1-3	0.387	0-0.128
1-4	0.234	0-0.111
1-5	0.211	0-0.116
2-3	0.196	0-0.064



2-4	0.293	0-0.086
2-5	0.315	0-0.097
3-4	0.645	0-0.045
3-5	0.599	0-0.066
4-5	0.726	0-0.04
<b>TASK – G1 vs G2</b>		
Models	p	95% CI
1-2	0.271	0-0.141
1-3	0.199	0-0.141
1-4	0.152	0-0.144
1-5	0.152	0-0.144
2-3	0.4	0-0.031
2-4	0.332	0-0.053
2-5	0.332	0-0.053
3-4	0.4	0-0.031
3-5	0.4	0-0.031
4-5	0.4	0-0.031
<b>TASK – G1 vs G3</b>		
Models	p	95% CI
1-2	0.34	0-0.178
1-3	0.299	0-0.187
1-4	0.552	0-0.204
1-5	0.346	0-0.18
2-3	0.4	0-0.058
2-4	0.742	0-0.147
2-5	0.828	0-0.112
3-4	0.607	0-0.141
3-5	0.816	0-0.104
4-5	0.105	0-0.315
<b>TASK – G2 vs G3</b>		
Models	p	95% CI
1-2	0.34	0-0.178
1-3	0.299	0-0.187
1-4	0.552	0-0.204
1-5	0.346	0-0.18
2-3	0.4	0-0.058
2-4	0.742	0-0.147
2-5	0.828	0-0.112
3-4	0.607	0-0.141
3-5	0.816	0-0.104
4-5	0.105	0-0.315

**Table S13.** Results of DeLong test with Bonferroni correction for each pair of models built for each classification task. p-values (p) for the difference in AUC and 95 % confidence interval (CI) for the difference in AUCs were reported.

## Section S6. Supplementary analysis for AJCC stage prediction.

In order to investigate the ability of T2 and DCE-MRI radiomic features in predicting AJCC stage, two classification tasks were investigated: stage I and II versus stage III and IV (I-II/III-IV classification task) and stage I versus II, III, and IV (I/II-III-IV classification task). Procedures described in the “Radiomic feature selection” and “Multivariable prediction model building and analysis” paragraphs were used. As reported in the following Tables, the step I and step II of feature selection returned respectively 51 and 10 (for I-II/III-IV), 53 and 8 (for I/II-III-IV). For I-II/III-IV classification task, the simplest multivariable model best prediction performances were reached by the second order model (AUC = 86%, sen = 65%, spec = 82%, acc = 77%). For I-II/III-IV classification task, the simplest multivariable model best prediction performances were reached by the second order model (AUC = 77%, sen = 75%, spec = 65%, acc = 72%). However, these

results must be interpreted with caution since there were multiple changes to the AJCC classification over the time period during which HCC specimens were collected (resulting in differences in the pathologic AJCC staging criteria across different editions). This constitutes a limitation in using AJCC stage as outcome. In particular, the AJCC TNM classifications used across cases in the TCGA-LIHC dataset range from the 5th through the 7th editions, resulting in inconsistency in the meaning of the pathologic T, N, and M categories across different patients examined during different time periods. These inconsistencies could also justify the differences in prediction performances obtained when the early stage group included stage I and II patients with respect to those obtained when the early stage group included only stage I patients.

#### S6.2 Prediction performances of multivariable logistic regression models.

ART_original_glcml_InverseVariance
ART_original_firstorder_Skewness
PORT_original_firstorder_10Percentile
PORT_original_firstorder_Kurtosis
PORT_original_firstorder_Skewness
TARD_original_firstorder_Kurtosis
T2_original_glcml_InverseVariance
T2_original_gldm_DependenceNonUniformityNormalized
T2_original_glszm_SizeZoneNonUniformityNormalized
T2_original_glszm_SmallAreaEmphasis

**Table S14.** 10 selected features after Wilcoxon rank-sum test for I-II/III-IV classification task.

SHAPE_original_shape_SurfaceVolumeRatio
ART_original_firstorder_Skewness
PORT_original_glszm_SizeZoneNonUniformityNormalized
PORT_original_firstorder_10Percentile
TARD_original_glszm_SizeZoneNonUniformityNormalized
TARD_original_firstorder_Skewness
T2_original_glszm_SizeZoneNonUniformityNormalized
T2_original_glszm_SmallAreaHighGrayLevelEmphasis

**Table S15.** 8 selected features after Wilcoxon rank-sum test for I/II-III-IV classification task.

#### S6.2 Prediction performances of multivariable logistic regression models.

Order	Features	AUC $\pm$ SE	SEN $\pm$ SE	SPEC $\pm$ SE	ACC $\pm$ SE	MCC $\pm$ SE
1	T2 glszm SZNUN	0.818 $\pm$ 0.004	0.517 $\pm$ 0.016	0.808 $\pm$ 0.006	0.744 $\pm$ 0.004	0.258 $\pm$ 0.019
2	T2 glszm SZNUN PORT firstorder 10Percentile	0.859 $\pm$ 0.005	0.646 $\pm$ 0.013	0.815 $\pm$ 0.005	0.775 $\pm$ 0.004	0.386 $\pm$ 0.015
3	T2 glszm SZNUN PORT firstorder 10Percentile PORT firstorder Kurtosis	0.857 $\pm$ 0.005	0.656 $\pm$ 0.014	0.795 $\pm$ 0.006	0.765 $\pm$ 0.004	0.381 $\pm$ 0.016
4	T2 glszm SZNUN PORT firstorder 10Percentile PORT firstorder Kurtosis T2 glszm SAE	0.82 $\pm$ 0.007	0.61 $\pm$ 0.015	0.789 $\pm$ 0.006	0.753 $\pm$ 0.005	0.328 $\pm$ 0.017
5	T2 glszm SZNUN PORT firstorder	0.788 $\pm$ 0.008	0.6 $\pm$ 0.016	0.779 $\pm$ 0.006	0.743 $\pm$ 0.005	0.283 $\pm$ 0.018

	10Percentile PORT firstorder Kurtosis T2 glszm SAE ART glcm IV					
--	--	--	--	--	--	--

**Table S16.** Results of multivariate analysis for I-II/III-IV classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; SZNUN = Size Zone Non Uniformity Normalized; SAE = Small Area Emphasis; IV = Inverse Variance.

Order	Features	AUC ± SE	SEN ± SE	SPEC ± SE	ACC ± SE	MCC ± SE
1	PORT firstorder 10Percentile	0.737 ± 0.006	0.709 ± 0.01	0.561 ± 0.012	0.655 ± 0.006	0.265 ± 0.017
2	TARD firstorder Skewness PORT firstorder 10Percentile	0.762 ± 0.006	0.746 ± 0.008	0.565 ± 0.012	0.683 ± 0.005	0.293 ± 0.016
3	T2 glszm SAHGLE TARD firstorder Skewness PORT firstorder 10Percentile	0.771 ± 0.007	0.752 ± 0.007	0.645 ± 0.012	0.715 ± 0.005	0.372 ± 0.016
4	T2 glszm SAHGLE TARD firstorder Skewness PORT firstorder 10Percentile SHAPE shape SVR	0.751 ± 0.007	0.725 ± 0.007	0.622 ± 0.012	0.691 ± 0.005	0.29 ± 0.018
5	T2 glszm SAHGLE TARD firstorder Skewness PORT firstorder 10Percentile SHAPE shape SVR ART firstorder Skewness	0.723 ± 0.008	0.725 ± 0.007	0.603 ± 0.013	0.687 ± 0.005	0.264 ± 0.019

**Table S17.** Results of multivariate analysis for I/II-III-IV classification task. For each model (from order 1 to 5), AUC, sensitivity, specificity and accuracy were reported with the standard error on a 95% confidence interval over all bootstrap sample. Abbreviations: T2 = features extracted from T2 images; ART = features extracted from arterial post-contrast phase of DCE-MRI; PORT = features extracted from portal post-contrast phase of DCE-MRI; SAHGLE = Small Area High Gray Level Emphasis; SVR = Surface Volume Ratio.

### S6.3 Equations for predictive logistic regression models.

$$g_{I-II/III-IV}(x_i) = 2.04 \times (\text{PORT ngtdm Complexity}) - 1.77 \times (\text{PORT glszm Large Area Low Gray Level Emphasis}) - 0.64$$

$$g_{I/II-III-IV}(x_i) = 1.24 \times (\text{T2 glszm SAHGLE}) + 1.16 \times (\text{TARD firstorder Skewness}) - 1.15 \times (\text{PORT firstorder 10Percentile}) - 0.52$$

### S6.4 Results of DeLong test with Bonferroni correction for each pair of models built for each classification task (I-II/III-IV and I/II-III-IV).

TASK – I+II vs III+IV		
Models	p	95% CI

1-2	0.122	0-0.101
1-3	0.118	0-0.110
1-4	0.203	0-0.105
1-5	0.145	0-0.113
2-3	0.773	0-0.0540
2-4	0.726	0-0.0450
2-5	0.884	0-0.0540
3-4	0.332	0-0.0320
3-5	0.779	0-0.0280
4-5	0.308	0-0.0380
<b>TASK – I vs II+III+IV</b>		
Models	p	95% CI
1-2	0.324	0-0.137
1-3	0.204	0-0.17
1-4	0.212	0-0.179
1-5	0.191	0-0.181
2-3	0.269	0-0.08
2-4	0.198	0-0.084
2-5	0.248	0-0.088
3-4	0.828	0-0.031
3-5	0.884	0-0.046
4-5	0.515	0-0.031